

Key Issue

Scientometrics, bibliometrics, altmetrics: some introductory advice for the lost and bemused

Metrics and their use and mis-use are very live issues for our communities. January 2013 sees the launch of the submissions system for the Research Excellence Framework (REF) 2014, the new system for assessing the quality of research in UK higher education institutions (HEIs). Bibliometric data such as citations will be used by the REF panels as part of their deliberation. While publications are still a key part of that analysis, panels are explicitly forbidden to consider the impact factor of the journal where those publications appear, when assessing the article's impact or importance: 'No sub-panel will make any use of journal impact factors, rankings, lists or the perceived standing of publishers in assessing the quality of research outputs'.¹

Alongside that, we see the rise of 'altmetrics' and 'scientometrics' – new ways of looking at the usefulness and impact of research outputs that take into account references outside of the journal article, including social media and news stories. The web makes these new metrics possible, and technology brings assessing impact and reach closer to real time.

Librarians, those working in research offices and publishers are no strangers to the use of metrics: for making journal collection decisions; showing return on investment of journal collections or funding of research; and assessing research outputs and impact institution. Often, however, we lack the time and resources to understand all the metrics available, let alone to gather the data and analyse it effectively to aid reporting and decision-making.

Every day seems to bring a new perspective on the 'best' way to assess impact, excellence and usage. There are many metrics, which all have their advantages, and their limitations. Where to begin? This article provides a brief introduction to some of the most commonly talked-about metrics. It also highlights some of the tools available to gather and analyse metrics. This is by no means a comprehensive survey of the metrics or services available.

Some article-level metrics

Citations

Citation is a key metric to understand, as many other metrics derive from citations.

A citation is a reference to a published work or an individual, in the references of a published work. This is done to acknowledge the source of information, or to substantiate a statement.

In this context, we are specifically referring to the citation of a journal article included in the references of another journal article. For journal articles, these are collated by a number of services, including Thomson Reuters' Web of Science, Elsevier's Scopus and Google Scholar.

Downloads

The number of times an article has been downloaded from a publisher's website or other source, such as an aggregator's database, or a publicly accessible repository like PubMed Central. A download can be viewing the abstract, the full-text HTML file or the PDF of an article. To ensure that online usage statistics are recorded and reported in a consistent way across the information industry, an international initiative, COUNTER (Counting Online Usage of Networked Electronic Resources) sets standards and provides a Code of Practice for information providers.



Grace Baynes
Head of Corporate
Communications
Nature Publishing
Group

312 Article downloads are often used to calculate 'cost per download' for journals (explained below). One potential limitation of this metric is that an article download gives no insight as to whether an article has been read, or used.

F1000 rating

Faculty of 1000 is a post-publication peer-review system, providing qualitative analysis and article rankings. 'Faculty members' recommend the best articles they have read in their specialist field, and give them a rating: 'Recommended', 'Must Read' or 'Exceptional'. This is then translated into a numerical score, the F1000 Article Factor. Faculty of 1000 is a proprietary subscription service from Science Navigation Group.

Social media mentions/links

An emerging set of metrics, this could include the number of track-backs to an article from blog posts; links from social networks such as Facebook, Twitter, Google+, Stumbleupon or Digg; or bookmarks in reference management tools like Mendeley and Readcube. The authors of *altmetrics: a manifesto* argue that such metrics are 'great for measuring impact in this diverse scholarly ecosystem'.² Tools such as ReaderMeter and Altmetric provide aggregation and visualization of these emerging 'scientometrics'.

Some journal-level metrics

Citations

Citations at journal level refer to the total of all citations to articles published in a journal over a given time period. Journal citations are usually reported in calendar years. These are collated by a number of services, including Thomson Reuters' *Journal Citation Report*, Elsevier's Scopus and Google Scholar.

Impact factor

An impact factor is a measure of how often the average article in a journal is cited in a given year. Impact factors apply to journals, not articles. The impact factor of a journal is calculated by:

$$\text{Impact factor for year } x = \frac{\text{Total citations in year } x}{\text{Total items published during years } x-1 \text{ and } x-2}$$

$$\text{So for example, a 2011 impact factor (published in June 2012) = } \frac{\text{Total citations in 2011}}{\text{Total items published in 2009 and 2010}}$$

The impact factor is a proprietary measure, calculated by Thomson Reuters. Impact factors are drawn from Web of Science data, and are published annually in the *Journal Citation Reports*. Impact factors have been criticized³ for being open to manipulation, being difficult to reproduce accurately, and for being inappropriately applied as a proxy measure of quality or impact for articles or researchers.

For further information on impact factors, see Jo Cross's chapter in UKSG's open access handbook, *The E-Resources Management Handbook* (<http://dx.doi.org/10.1629/9552448-0-3.17.1>)

Cost per download

Cost per download is a way of estimating the per-article cost for a journal. The subscription cost of the journal is divided by the number of article downloads in the subscription period, most often a year. This metric can be used to compare journals from different publishers that publish different quantities and that have different prices. It can be used as an indicator of return on investment for an institution, or by a publisher to demonstrate value for money. However, a download does not necessarily indicate that the article was used, only that it was accessed. Whether the article was used by a student for undergraduate study, or a researcher to inform research projects is unknown.

Journal usage factor

The journal usage factor, a UKSG initiative, is now run by COUNTER. It aims to look at use (in terms of downloads) of journals in the same way that the impact factor looks at citations.

$$\text{Journal usage factor} = \frac{\text{Total usage over period } x \text{ of items published during period } y}{\text{Total items published online during period } y}$$

The project is still at an exploratory stage and the metric has not yet been widely adopted.

Eigenfactor

The Eigenfactor aims to give a numerical indicator of the overall contribution of the journal to the literature. It is based on citations, and uses the Thomson Reuters Web of Science citation counts. The Eigenfactor uses an algorithm to apply a weighting to citations, to take into account the number of citations within the journal, and where citations to the journal are coming from. A good parallel is Google's 'page rank', which takes into account the number of links into and out of a page, and where the links come from. Journals that publish more citable articles will, all things being equal, have a higher Eigenfactor score. The Eigenfactor has the potential to be very useful, as it provides a richer analysis of citations than a straight count, but it is quite complex to understand and explain. The creators of Eigenfactors make their methodology publicly available.

h-index

The h-index is another potentially very useful metric, which is also quite difficult to succinctly explain. The h-index was developed by Jorge E Hirsch 'to quantify a researcher's output'. The h-index aims to measure both productivity and impact, by counting the number of papers and the number of citations to those papers. It is defined as: 'the number of papers with citation number $\geq h$ '.⁴ For example, a researcher with an h-index of 5 has published five papers that have each been cited five or more times. The researcher in question may have published an additional ten papers that have all been cited four times or less, but these do not count towards the h-index.

Recently, the h-index has been applied to journals; it can also be extended to groups of researchers, and variants of it form the basis of the Google Scholar Metrics.

More metrics ...

There are many more metrics available and in use than are described here. There is also the m-index, the c-index, the g-index, the e-index ...

Metrics and analytics tools

Without the time and tools to analyse, collecting metrics is of little use. Technology offers hope for this, and there are a growing number of tools available. Some examples include:

ResearcherID

This is a Thomson Reuters service. Citation metrics for individual researchers are available, amongst other services and information, based on Web of Science data (web-based).

Altmetric

Altmetric captures social media mentions, blog posts, news stories and other pieces of content that mention scholarly articles. Digests, scores and displays in visual format, with ability to dive into data. Looks at individual articles or journals. Supported by Digital Science (web-based, subscription access).

Journal Usage Statistics Portal (JUSP)

Available to all HEIs in the UK and supported by JISC, this portal harvests COUNTER-compliant usage statistics, enabling libraries to, for example, compare usage across journals and publishers packages, view the highest-accessed journals and generate reports (web-based, free to HEIs in the UK).

SciVal Analytics

These are bespoke research projects and reports from Elsevier, for 'measuring and managing research performance' (on-demand/bespoke, paid-for service).

Symplectic Elements

Installed in university systems, Symplectic Elements can ingest data from CrossRef, Web of Science and others. Reports available on publications, citations, h-index and altmetrics. Supported by Digital Science (installation, paid-for service).

IRUS

This new service forms part of the JISC-funded repository and infrastructure service, UK RepositoryNet+ (RepNet). It aims to enable UK institutional repositories to access and share comprehensive and comparable usage statistics using the COUNTER standard. The service will collect usage data from participating repositories, process the data into COUNTER-compliant statistics and then present statistics back to originating repositories.

For further information about this project, see Paul Needham and Graham Stone's article <http://dx.doi.org/10.1629/2048-7754.25.3.262> in this issue or the IRUS website (<http://www.irus.mimas.ac.uk/>).

The importance of context

Metrics are only useful if you know what they mean, what their limitations are, and if you consider them in context. Most journal-level metrics are only useful as a comparator to other journals within the same field. The same is largely true of article-level metrics – it is not meaningful to compare the citation level of a genomics paper with that of a high-energy physics paper or a clinical case study. The research communities in question do not all publish or cite papers with the same frequency.

"Metrics are only useful if you know what they mean, what their limitations are, and if you consider them in context."

Use of metrics in decision-making

Some institutions are looking in depth at journal-by-journal metrics, some are looking at a publisher level. What the librarians and information managers look at depends on whether they purchase journal by journal, or in 'bundles' or 'big deals', and on the time and resources available to perform such analysis. For some, particularly those facing budget cuts, it is the price tag that becomes the deciding factor. For others, it is what the faculty require that trumps all, even if the statistics show that 'must-have' journals are rarely used.

315 As in the research fields we all serve, statistics are useful and potentially powerful when applied correctly. In order to make use of them, we need to know what questions we are trying to answer, understand what data and metrics will provide the answers we seek, understand the limitations of the metric(s) we have chosen to apply, have the time and resources to gather and analyse them, and the willingness and ability to make changes based on our findings.

Blogs for suggested further reading:

<http://blogs.warwick.ac.uk/libresearch/tag/bibliometrics/>

<http://sharmanedit.wordpress.com/>

References

1. REF2014:
<http://www.ref.ac.uk/faq/all/> (accessed 10 September 2012).
2. Altmetrics.org:
<http://altmetrics.org/manifesto/> (accessed 10 September 2012).
3. Brumback, Roger A, Impact Factor Wars: Episode V--The Empire Strikes Back, *Journal of Child Neurology*, (2009) 24, 3, 260–262.
<http://dx.doi.org/10.1177/0883073808331366> (accessed 10 September 2012).
4. Hirsch, J E, An index to quantify an individual's scientific research output, *PNAS* (2005) 102, 46 16569–16572:
<http://dx.doi.org/10.1073/pnas.0507655102> (accessed 10 September 2012).

Disclosure: Digital Science and Nature Publishing Group are both part of Macmillan Publishers Limited. Grace Baynes is an employee of Macmillan Publishers Limited.

The author claims to be no expert on metrics, but an interested member of the publishing community, and offers the above on that basis.

Key Issue © Grace Baynes

E-mail: g.baynes@nature.com

To cite this Key Issue:

Baynes, G, Scientometrics, bibliometrics, altmetrics: some introductory advice for the lost and bemused, *Insights*, 2012, 25(3), 311–315, doi: 10.1629/2048-7754.25.3.311

To link to this article:

<http://dx.doi.org/10.1629/2048-7754.25.3.311>